

Genomic binding by the *Drosophila* Myc, Max, Mad/Mnt transcription factor network

Amir Orian,¹ Bas van Steensel,³ Jeffrey Delrow,¹ Harmen J. Bussemaker,⁴ Ling Li,¹ Tomoyuki Sawado,¹ Eleanor Williams,² Lenora W.M. Loo,¹ Shaun M. Cowley,¹ Cynthia Yost,¹ Sarah Pierce,¹ Bruce A. Edgar,¹ Susan M. Parkhurst,¹ and Robert N. Eisenman^{1,5}

¹Division of Basic Sciences and ²Division of Human Biology, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, USA; ³Netherlands Cancer Institute, 1066 CX Amsterdam, The Netherlands; ⁴Department of Biological Sciences and Center for Computational Biology and Bioinformatics, Columbia University, New York, New York 10027, USA

The Myc/Max/Mad transcription factor network is critically involved in cell behavior; however, there is relatively little information on its genomic binding sites. We have employed the DamID method to carry out global genomic mapping of the *Drosophila* Myc, Max, and Mad/Mnt proteins. Each protein was tethered to *Escherichia coli* DNA adenine-methyltransferase (Dam) permitting methylation proximal to in vivo binding sites in Kc cells. Microarray analyses of methylated DNA fragments reveals binding to multiple loci on all major *Drosophila* chromosomes. This approach also reveals dynamic interactions among network members as we find that increased levels of dMax influence the extent of dMyc, but not dMnt, binding. Computer analysis using the REDUCE algorithm demonstrates that binding regions correlate with the presence of E-boxes, CG repeats, and other sequence motifs. The surprisingly large number of directly bound loci (~15% of coding regions) suggests that the network interacts widely with the genome. Furthermore, we employ microarray expression analysis to demonstrate that hundreds of DamID-binding loci correspond to genes whose expression is directly regulated by dMyc in larvae. These results suggest that a fundamental aspect of Max network function involves widespread binding and regulation of gene expression.

[Keywords: *myc*; *mad*; *Drosophila*; target genes; transcription]

Supplemental material is available at <http://parma.fhcrc.org/AOryan>.

Received December 12, 2002; revised version accepted February 12, 2003.

The Myc, Mad, Max network comprises a group of widely expressed transcription factors, which function in cell proliferation and differentiation. This network includes members of the Myc and Mad families, the Mad-related protein Mnt, and Mga. All of these proteins possess basic helix-loop-helix zipper domains (bHLHZ), which mediate dimerization with the small bHLHZ protein Max, thereby forming heterodimers capable of recognizing the E-box sequence CACGTG (Grandori et al. 2000). Although Max homodimers lack transcriptional activity, recent evidence suggests that the transcriptional activities of Max heterodimers derive from their ability to recruit chromatin-modifying complexes to DNA (Amati et al. 2001; Eisenman 2001). For example, Mad family proteins directly bind the mSin3-histone deacetylase corepressor complex permitting Mad-Max heterodimers to target histone deacetylation at E-box bind-

ing sites (Laherty et al. 1997; Bouchard et al. 2001). Myc proteins associate with the TRRAP coactivator, which in turn binds the histone acetyltransferase, GCN5, permitting Myc-Max dimers to direct acetylation of histones at E-box binding sites (McMahon et al. 2000; Bouchard et al. 2001; Frank et al. 2001). Furthermore, Myc is known to bind and inhibit the function of the transcription factor Miz-1, leading to E-box-independent repression of Miz-1 activated targets such as p15^{INK4b} and other cyclin-dependent kinase inhibitors (Staller et al. 2001; Ziegelbauer et al. 2001; Seoane et al. 2002).

The varied activities of Max network proteins are manifested during proliferation and differentiation. Mad family proteins are generally induced during terminal differentiation and act to limit cell proliferation (Zhou and Hurlin 2001). Myc proteins are induced in response to a large number of growth factors and cytokines and may serve to integrate these external signals to sustain growth and proliferation. Targeted deletion of either *N-myc* or *c-myc* leads to early embryonic lethality in the mouse (Stanton et al. 1992; Davis et al. 1993) and recent studies demonstrate that *myc* gene function is essential

⁵Corresponding author.

E-MAIL eisenman@fhcrc.org; FAX (206) 667-6522.

Article published online ahead of print. Article and publication date are at <http://www.genesdev.org/cgi/doi/10.1101/gad.1066903>.

for hematopoiesis and organogenesis (de Alboran et al. 2001; Douglas et al. 2001; Trumpp et al. 2001; Knoepfler et al. 2002). Deregulated overexpression of *myc* genes leads to malignant transformation, genetic instability, and apoptosis. In contrast to Myc and Mad, Max is a stable, constitutively expressed protein acting primarily as an obligate dimerization partner permitting Myc and Mad proteins to associate with DNA.

The Max network is highly conserved and *Drosophila* orthologs of vertebrate Myc and Max (dMyc and dMax) have been identified and characterized previously (Gallant et al. 1996; Schreiber-Agus et al. 1997; Johnston et al. 1999). More recently, a fly ortholog of mammalian Mad/Mnt was identified (termed dMnt; Bourbon et al. 2002; L.W.M. Loo and R.N. Eisenman, in prep.). Whereas vertebrates possess families of *myc* and *mad* genes, *Drosophila myc*, *max*, and *mnt* have no paralogs. Both dMyc and dMnt form heterodimers with dMax that specifically bind CACGTG. Whereas dMyc–dMax heterodimers activate transcription, dMnt–dMax associates with *Drosophila* Sin3 and represses transcription in an E-box-dependent manner (Gallant et al. 1996; L.W.M. Loo and R.N. Eisenman, in prep.). Moreover, similar to mammalian *myc*, *dmyc* is capable of cotransforming primary mammalian cells and rescuing the proliferation defect in *c-myc* null fibroblasts (Schreiber-Agus et al. 1997; Trumpp et al. 2001). Furthermore, both vertebrate and *Drosophila myc* regulate cell growth. The molecular and biological similarities between *Drosophila* and vertebrate Myc, Max, Mad, Mnt proteins coupled with the availability of *Drosophila* genetic tools and the complete genome sequence (Rubin et al. 2000) make *Drosophila* an attractive system to carry out a systematic analysis of genomic binding by the network. In this study, we have employed the recently devised DamID method (van Steensel and Henikoff 2000; van Steensel et al. 2001) in which a bacterial DNA methylase, fused to dMax network transcription factors, is used to “mark” DNA-binding sites in living cells.

Results

To identify genomic binding regions for Max network transcription factors we prepared fusion proteins consisting of the bacterial DNA adenine methylase (Dam) linked to the C termini of full-length dMax, dMnt, and dMyc as well as to the N terminus of dMyc. In each case, the transcription factor sequence was separated from Dam by a 9E10 (human Myc tag) peptide, permitting detection of the fusion protein with 9E10 antibody (see Materials and Methods for details). All of these fusion proteins were found to localize to the cell nucleus and activated (dMyc–Dam and Dam–dMyc) or repressed (dMnt–Dam) transcription from a synthetic reporter gene similar to their vertebrate orthologs (Grandori et al. 2000). In contrast, the tagged Dam protein alone, or a Myc mutant (amino acids 1–624) lacking the bHLHZ domain, had no effect on E-box-dependent transcription (data not shown; see Materials and Methods). These results indicate that fusion with Dam does not impair the

functions of dMyc and dMnt. In order to ensure low levels of expression, the chimeric proteins were expressed separately in *Drosophila Kc* cells under control of the heat-shock promoter, but in the absence of heat-shock (van Steensel and Henikoff 2000). At 24 h posttransfection, genomic DNA was extracted. Subsequently, 0.1–2-kb fragments generated by digestion with *DpnI* (which cuts only at G^{m6}ATC) were purified by sucrose gradient centrifugation, labeled with the fluorochromes Cy5 (chimeric proteins) and Cy3 (Dam alone, serving as a reference for nonspecific binding/accessibility) and cohybridized to a *Drosophila* cDNA array that contains 6255 cDNAs and ESTs, representing roughly half of *Drosophila* coding sequences. Targeted sequences were identified based on the Cy5:Cy3 fluorescence ratio (van Steensel and Henikoff 2000; van Steensel et al. 2001; see Materials and Methods). The binding data for each protein on the entire array (chromatin profile) were generated as described previously (van Steensel et al. 2001). A set of statistical tests, as well as a “self on self” experimental set (dMnt–Dam vs. dMnt–Dam), were used to establish statistically significant targets and assess experimental noise (see Materials and Methods; Supplemental Material S3 for access to raw DamID binding data).

Comparative chromatin profiling of Max network proteins

Having established chromatin profiles and a statistically significant set of binding targets for each of the network proteins, we compared the profiles derived from experimental sets for each of the network proteins to examine binding overlap among the different members. Figure 1A is a scatter plot comparison between the chromatin profiles of dMnt and dMax in which we compare separate binding data sets for each protein. These profiles show a high degree of similarity (correlation coefficient, $r = 0.49$) with both proteins displaying statistically significant binding to the same 258 genes (Fig. 1A, purple points). Many genes are not bound by either protein (Fig. 1A, gray points). We also found that increasing dMax levels had no effect on dMnt genomic binding (data not shown). The Venn diagrams in Figure 2 depict the number of overlapping and nonoverlapping genes. Overlapping genes represent 60% of the total genes occupied by dMnt and 40% of dMax bound genes (Fig. 2A). DamID analysis of binding by GAF, a transcription factor unrelated to dMax network proteins (van Steensel et al. 2003), shows only minimal overlap with dMnt bound loci ($r = 0.11$; Fig. 2F), further supporting the idea that the overlap detected among dMax network proteins is specific.

A similar comparison of data sets derived from dMax and dMyc produced a surprising result (Fig. 1B). First, dMyc was recruited to a considerably smaller number of genes (22) than either dMax or dMnt (646 and 429 respectively; see Fig. 2B). Second, the dMax binding profile displayed little overlap with binding by either the C-terminal or N-terminal fusions of dMyc to Dam (Fig. 2B; data not shown). We asked whether dMax might be lim-

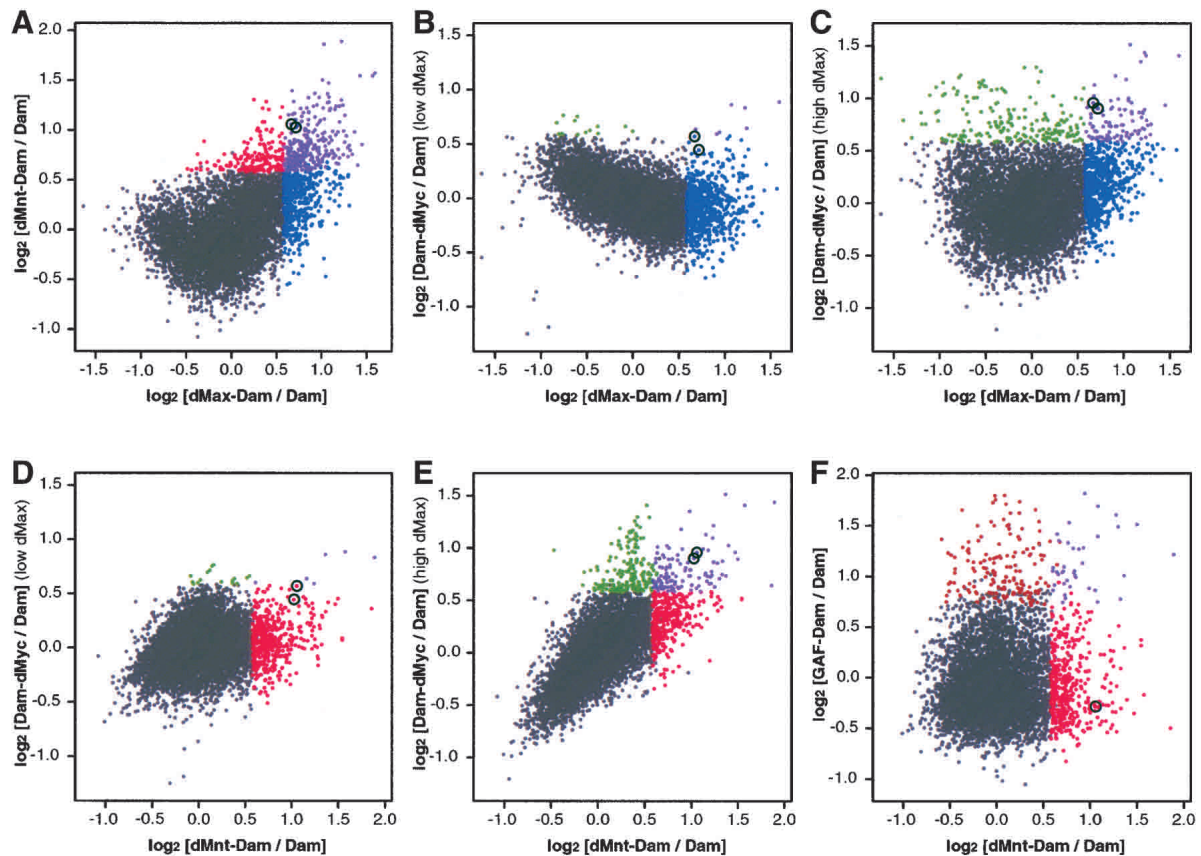


Figure 1. Comparison of chromatin profiles for the *Drosophila* Max network. Scatter plot comparisons of chromatin binding profile data sets of the dMax network proteins derived from the entire cDNA array. The average binding is presented as Cy5: Cy3 log ratios between dMax–Dam:Dam vs. dMnt–Dam:Dam, $r = 0.49$ (A); dMax–Dam:Dam vs. dMyc–Dam:Dam with dMax expressed at low levels, $r = -0.43$ (B); dMax–Dam:Dam vs. dMyc–Dam:Dam with dMax expressed at high levels, $r = 0.17$ (C); dMnt–Dam:Dam vs. dMyc–Dam:Dam where dMyc–Dam is coexpressed with low levels of dMax, $r = 0.26$ (D); dMnt–Dam:Dam vs. dMyc–Dam:Dam where dMyc–Dam is coexpressed with high dMax levels, $r = 0.71$ (E); dMnt–Dam:Dam vs. dGAF–Dam, $r = 0.11$ (F). Ratios were calculated using a full set of experiments as described in Materials and Methods. Pearson correlation coefficients (r) were calculated for each set. Statistically significant binding targets are indicated by colored solid dots. Blue, dMax–Dam; green, dMyc–Dam; red, dMnt–Dam; brown, dGAF–Dam; purple, overlapping targets. The *bicaudal* (*bic*) gene is marked by open black circle and is printed twice on the array.

iting for dMyc binding under our experimental conditions and therefore we expressed relatively higher levels of nonfused dMax along with the N-Dam–dMyc fusion (see Materials and Methods). As shown in Figures 1C and 2C, dMyc recruitment to targets and the binding profile overlap between dMyc and dMax were dramatically enhanced when dMax levels were augmented. Under these conditions the dMyc fusion associated with 287 targets, 96 shared with dMax (33.4%) and 73 (25.4%) shared with both dMnt and dMax (cf. Fig. 1B,C and the Venn diagrams in Fig. 2B,C,F). Augmented expression of dMax also resulted in increased overlap between dMyc and dMnt binding profiles, resulting in an overlap of 113 genes (39%; Figs. 1D,E, 2D–F). Myc recruitment to the majority of these target genes depends on functional dMax as the vast majority of the targets were not bound by dMyc when dMax lacking its leucine zipper domain (residues 91–119 deleted) was coexpressed (data not shown). Because the zipper region is essential for Max's

ability to heterodimerize this result demonstrates that association between Myc and Max proteins is most likely responsible for the increased overlap between identified target genes. Nonetheless, although dMax levels were limiting for Dam–dMyc binding, a substantial number of dMyc targets did not overlap with either dMax or dMnt in the presence of either low or high dMax levels (17 and 152 unique Myc targets respectively; Fig. 2B,F; see Discussion). These findings suggest that dMax is limiting for dMyc binding to certain targets in *Kc* cells. Furthermore, the fact that dMax binds a subset of sites distinct from those bound by dMnt and dMyc implies that dMax heterodimerizes with an as-yet-unidentified factor (see Discussion).

Distribution of dMax network binding loci

Of 6255 genes represented on the array, 968 (15.4%) were found to associate with one or more dMax network pro-

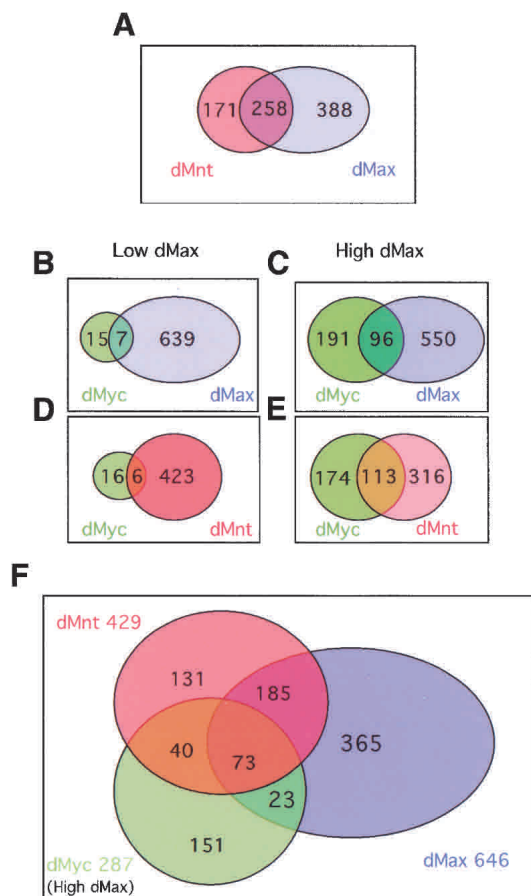


Figure 2. Venn diagrams depicting the number of shared genes within the dMax network. Criteria for calculating statistically significant targets are described in Materials and Methods. (A) Comparison of dMnt and dMax targets. (B,D) dMyc data sets generated in the presence of low dMax levels. (C,E,F) dMyc data sets generated in the presence of high dMax levels.

teins. The distribution and overlap is depicted in the Venn diagram in Figure 2F. Previous work showed that significant and specific methylation by tethered Dam occurs within 1.5–2 kb of the DNA-binding site (van Steensel and Henikoff 2000) permitting us to link binding regions detected on the array with a map of *Drosophila* chromosomes (Fig. 3A; see Discussion). Extensive binding of all three proteins was observed on the four major *Drosophila* chromosomes (Fig. 3A); however, repetitive DNA elements displayed relatively few binding sites despite the presence of E-boxes (e.g., Tirant; Fig. 3B). Even the low level of binding to repeats was not reproducible on identical repeats spotted multiple times on the array (e.g., the 297 and 412 repeats) and may therefore represent false positive hits. A previous DamID analysis showed that many of these repetitive elements are bound by *Drosophila* HP1 (heterochromatin protein 1), suggesting the lack of dMax network protein binding is likely to be specific, and not attributable to technical limitations or a generalized exclusion of chromatin-binding factors from this region (van Steensel et al.

2001). Therefore, although dMax network protein binding to chromatin is widespread, it appears to exclude specific regions, at least some of which are silenced regions associated with heterochromatin.

Binding regions correlate with E-box sequences as well as other motifs

The E-box sequence CACGTG and several related sequences are binding sequences for vertebrate and *Drosophila* Myc, Max, and Mad/Mnt proteins (Blackwell et al. 1990, 1993; Blackwood and Eisenman 1991; Prendergast and Ziff 1991; Solomon et al. 1993; Gallant et al. 1996; James and Eisenman 2002; L.W.M. Loo and R.N. Eisenman, in prep.). To determine whether such sites are represented in our binding regions and to substantiate our in vivo-binding results, we applied an unbiased bioinformatics method, the REDUCE algorithm. REDUCE identifies putative regulatory sequences of 8 bp or less whose presence correlates with binding (Bussemaker et al. 2001; for review, see Li 2002). The REDUCE algorithm employs the binding profiles of the entire array without any preclustering or cut-off data restrictions. The readout of the REDUCE search is a list of distinct motifs with their statistical significance and (positive) regression coefficients quantifying their contribution to binding to the locus in which they occur. As shown in Table 1, binding sites for the canonical E-box CACGTG display significant correlation with the dMnt binding log-ratio ($p < 10^{-12}$; Table 1A). However, the CACGTG correlated with dMyc binding only in the presence of high levels of coexpressed dMax ($p < 10^{-12}$; Table 1C). A high correlation with short CG repeats may reflect a previously noted preference for binding by Myc and Mad to E-boxes that contain CG in flanking regions (Prendergast et al. 1991; Solomon et al. 1993; James and Eisenman 2002). When exogenous dMax levels are low, only AT-rich motifs correlate with binding (see Supplemental Material S5).

Interestingly, REDUCE analysis also identified other sequence motifs that correlate with binding. Because of their high degree of association with Max network binding sites these motifs may serve as binding sites for factors that cooperate with dMyc and dMnt. One such putative binding site is the palindromic sequence TATC GATA (DNA replication element, DRE), a sequence reported near genes involved in cell proliferation and growth (Hirose et al. 1993, 2001; Jasper et al. 2002; Table 1, bold). The DRE was shown to be a consensus binding element for several factors: DREF (DNA replication element factor, a factor associated with the TRF2 complex; Hochheimer et al. 2002), Cut (a repression component within the Notch pathway; Jackson and Blochinger 1997; Nepveu 2001), and BEAF32 (a boundary element associated binding factor; Zhao et al. 1995). Interestingly, REDUCE analysis demonstrates that the DRE sequence was the most significant motif among dMax targets (Table 1), suggesting that dMax is involved in as yet unidentified complexes (see Discussion).

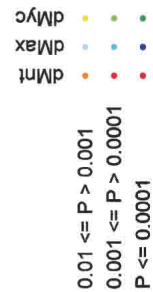


Figure 3. Map of dMax network binding across *Drosophila* chromosomes and repetitive elements. Annotation of binding targets was performed using the *Drosophila* gene collection subset (Rubin et al. 2000). Individual genes are depicted as vertical gray bars. dMyc, dMnt, and dMax binding sites are denoted by green, red, and blue solid circles, respectively. Color intensity is directly linked to binding P values as depicted at the bottom. (A) dMax network binding to the four *Drosophila* chromosomes. A partial list of genes found to be orthologs of direct targets of c-Myc (Fernandez et al. 2003) are indicated by black triangles and gene names. dMyc-regulated genes in the larval expression study that were detected in the the DamID analysis are denoted by purple (activated) and pink (repressed) arrows. (B) dMax network binding to repetitive elements.

Table 1. Highest scoring motifs detected by REDUCE analysis

Motif	R ²	P value	F	Matches	Loci	Consensus
A dMnt						
CGCG	0.051	0.0E + 00	0.01	66,217	4366	CG-repeat
GCGC	0.048	0.0E + 00	0.01	97,925	4367	CG-repeat
CGCGC	0.047	0.0E + 00	0.02	16,661	4121	CG-repeat
GCGCG	0.042	0.0E + 00	0.02	17,392	4129	CG-repeat
TATCGATA	0.026	0.0E + 00	0.06	1618	1258	TATCGATA
ATCGATA	0.024	0.0E + 00	0.04	3863	2367	TATCGATA
TCGATA	0.015	1.0E – 12	0.02	8262	3541	TATCGATA
TATCGAT	0.015	5.0E – 12	0.03	3765	2369	TATCGATA
GGTCACAC	0.024	0.0E + 00	0.09	788	706	GGTCACACT
GTCACACT	0.017	0.0E + 00	0.08	691	633	GGTCACACT
CACGTG	0.019	0.0E + 00	0.03	4214	2558	CACGTG
GCACGTG	0.016	0.0E + 00	0.05	1370	1139	CACGTG
GCACGTGT	0.012	9.8E – 09	0.10	319	301	CACGTG
B dMax						
CGCGC	0.025	0.0E + 00	0.02	16,401	4058	CG-repeat
CGCG	0.022	0.0E + 00	0.01	65,098	4301	CG-repeat
GCGCG	0.020	0.0E + 00	0.02	17,086	4069	CG-repeat
GCGC	0.014	2.0E – 12	0.00	96,364	4302	CG-repeat
TATCGATA	0.024	0.0E + 00	0.07	1600	1247	TATCGATA
ATCGATA	0.018	0.0E + 00	0.04	3797	2330	TATCGATA
TATCGAT	0.016	2.0E – 12	0.04	3705	2334	TATCGATA
GGTCACAC	0.013	2.0E – 09	0.08	776	693	GGTCACACT
GTCACACT	0.009	1.1E – 05	0.07	681	624	GGTCACACT
C dMyc (high dMax)						
CACGTG	0.032	0.0E + 00	0.03	4203	2555	CACGTG
ACGTG	0.024	0.0E + 00	0.01	18,323	4228	CACGTG
CACGT	0.022	0.0E + 00	0.01	17,082	4221	CACGTG
GCACGTG	0.021	0.0E + 00	0.05	1365	1134	CACGTG
ATCGATA	0.022	0.0E + 00	0.03	3853	2362	TATCGATA
TCGATA	0.022	0.0E + 00	0.02	8255	3535	TATCGATA
TATCGATA	0.020	0.0E + 00	0.04	1616	1256	TATCGATA
ATCGAT	0.014	2.2E – 11	0.01	13,336	4037	TATCGATA
TATCGAT	0.013	2.1E – 10	0.02	3751	2362	TATCGATA
CGCGC	0.027	0.0E + 00	0.01	16,646	4116	CG-repeat
CGCG	0.026	0.0E + 00	0.00	66,118	4361	CG-repeat
GCGC	0.024	0.0E + 00	0.00	97,752	4362	CG-repeat
GCGCG	0.019	0.0E + 00	0.01	17,348	4124	CG-repeat

For each spot on the microarray the Berkeley *Drosophila* Genome Sequence Release 2 was used to generate an associated DNA region comprising the sequence of the array probe plus 2 kb flanking region. The Matches column indicates the total number of matches to the motif in the combined sequences. The Loci column indicates the number of sequences with at least one match. F represents the regression coefficient for a given motif. R² equals the square of the Pearson correlation between the binding log ratio and the motif count (see Materials and Methods). All oligonucleotides of up to 8 bp in length were tested in parallel. Only the most significant motifs ($p < 10^{-4}$) with a positive correlation are shown. E-box and DRE sequences are depicted in bold. Complete REDUCE results available in Supplemental Material S5.

Binding regions are proximal to genes implicated in Max network functions as well as genes regulated by dMyc

We sorted dMax network binding sites into functional groups using the *Drosophila* Gene Ontology database (GO; <http://www.godatabase.org>), a gene annotation system (Ashburner et al. 2000). Table 2 is a partial list of annotated gene targets. Our screen reveals both novel and previously described pathways and target genes associated with dMax network function (see Discussion). The complete binding profiles and partial annotation can be found as Supplemental Material (S1, S2).

We next asked whether the binding loci defined by

DamID in *Kc* cells could be identified as dMyc-responsive genes in the rapidly growing third instar larvae. Using the heat-shock Flp/Gal4 method (Pignoni and Zipursky 1997; Neufeld et al. 1998), we have generated transgenic flies expressing a marker GFP and dMyc, or GFP alone, both under UAS control. Total RNAs collected from either Myc-expressing or control larvae 7 and 14 h after heat-shock induction were Cy3/Cy5 labeled and hybridized to the same *Drosophila* microarrays used for the DamID experiments (see Materials and Methods). Statistically significant mRNA expression changes were defined for 845 genes: 544 genes activated and 301 repressed (see Supplemental Material S1, S4). Comparison of the dMax network DamID binding profile with the

Table 2. Partial list of the dMax network target loci

Biological pathway/ molecular function	dMnt targets	dMax targets	dMyc targets
Actin/cytoskeleton/cell migration	Dmn, nup44A, <i>Tm1</i> , Sop2, CG6546, CG25C, Dlc90F, nop5, bif	nup44A, <i>Tm1</i> , Arp11, bif, CG15669, puc, Klp3A	<i>Tm1</i>
Cell cycle regulators	Myt1, ial, slbp	ial, cdk4, cdc2, cdc2c, Rbf, CycJ, CycC, CycD, stg	cdk4, CycA, CycB, CycB3, Rbf
Cell death	Dredd	Dredd, Dcp-1, Ice, debcl	Reaper L
Channels/transporters	CG10444, <i>icln</i>	CG10444, Nrv1, Nhe1, <i>icln</i>	rpK, <i>icln</i>
Chromatin	<i>HP1b</i> , HmgD, Df31, Nap1, Caf1	<i>HP1b</i> , HmgD, Caf1	<i>HP1b</i> , fs (1) Ya
Immunity	FK506bp1, Dredd, <i>Rel</i> , PGRP-LE	FK506 bp1, Myd88, Rel, PGRP-LE, CG10535, CG3829	<i>Rel</i>
Mitochondria biogenesis, structure, function	<i>Tim10</i> , Tim9, mRSp30, mRp19, <i>mRpS7</i> , TFAM, Ferrochelatase	<i>Tim10</i> , mRpS30, mRp19, <i>mRpS7</i> , TFAM, colt	<i>Tim10</i> , mRpL10, <i>mRpS7</i> , TFAM, CG3476, mge
Nucleolus	CG12909, CG8939, Surf6	CG1135	CG1135, CG12909, Rpp30, Surf6
Peptidyl prolyl cis-trans isomerase	Fk560bp1	CG3511, FK506bp1, FKBP59	CG5482, cyp33
Protein synthesis/Ribosome	Aast-gln, eIF6, mRpL19, mRpS30, <i>mRpS7</i> , <i>RpL1</i> , RpS5, RpS9, Nmd3, Surf6, CG1475	bonsai, cg1475, mRpL19, <i>mRpS7</i> , <i>RpL1</i> , eIF6, pelo	<i>RpL1</i> , RpL19, pelo, <i>mRpS7</i> , mRpL1, RpS12, RpS13, RpS9, Surf6
Replication/DNA repair	<i>RnrS</i> , CG8142, pms2	<i>RnrS</i> , Rfc40, Mcm2, Mcm6, SPT4, Orc1, mus209, CG8142	<i>RnrS</i> , UbcD6
RNA binding	CG11738, CG12909, CG14230, CG5589, CG8636, CG8862, mod, Nmd3, Rbp1, SC35, SF2, Slbp, tra2, U2af38	BcDNA: GH01073, CG11738, CG14230, CG1704, CG5589, CG8862, Hrb98DE, mxc, Nmd3, qkr58E-1, Rbp1, SC35, snf, Srp54	CG10214, CG11738, CG12909, CG8862, cyp33, mod
RNA methyltransferase	CG11837, CG5220, CG7319, CG7818	CG5220, CG7319	CG1837, CG5220
Signaling	<i>aay</i> , CG1815, CG4527, CG6805, LD08534	loco, pak3, spz, grk, PP2-AB, puc, stg, cg66805, LD08534, <i>aay</i>	<i>aay</i> , LD08534
Small GTPases GTP binding proteins	Rab14, RhoL	Rab17, Ran, Rap21, Ras85D, RhoL, Arf84F	Rab11
Spliceosome	CG10754, CG4980, CG6876, SC35, SF1, SF2, SmB, U2af38	SC35, Snf, Srp54	
Synaptic vesicles traffic	Syx13, Syx16, Syx17, Syx18, Syndapin, usnp, Rop, KdelR	Syx17, Syx18, Arf84D, Syndapin, Rop, KdelR	Syx13, CG8862
Transcription factors	<i>bic</i> , Max, DREF, Pan, Ell, TFAM, Rel, Ira, Med6	<i>bic</i> , DREF, bigmax, myb, Xbp1, Rpl133, Rbf, TFAM	<i>bic</i> , bigmax, E2F2, Rel, Rbf, TFAM
Transcription initiation and elongation	Taf55, Tfb2, <i>Cyp1</i>	Taf60-2, <i>Cyp1</i>	Taf60-2, <i>Cyp1</i>
Transferase transferring glycosyl groups	CG18012, CG4802, CG5537, CG6437, CG9249	CG5220, CG7319, CG9249	CG3434, CG4802, CG5537, CG9249
Translation factors	EfTuM, Adam, CG5705	EfTuM, Adam, eif6, eif4A	Adam
Ub-proteasome	CG5384, CG5505, <i>smt3</i> , Ufd1-like, Rpn5, Rpn7, Pros25, Pros26	CG5384, <i>smt3</i> , Ufd1-like, Rpn5, Rpn7, Pros45, Tbp1, Uba1	UbcD6, <i>smt3</i>

Standardized gene annotation was applied from the *Drosophila* Genome Ontology database (release February 2002). Only annotations in the biological process and molecular function categories with three or more targets are listed. A list of the network target genes can be found in Supplemental Material S2. Shared targets between all three proteins are in italic.

gene expression profile indicates a high degree of overlap (Fig. 3, purple and pink arrows; Table 3). The highest

degree of correlation between binding and expression was observed for loci that are bound by both dMyc and

Table 3. Comparison of *DamID* binding loci with dMyc-induced larval gene expression targets

A Loci bound by	B No. of binding loci identified by <i>DamID</i>	C No. of dMyc regulated genes also identified by <i>DamID</i>	D Percent of overlap	E Activated genes	F Repressed genes
dMyc + dMax + dMnt	73	35	47.9%	34	1
dMyc + dMnt	40	24	60.0%	24	0
dMyc + dMax	23	6	26.0%	5	1
dMnt + dMax	185	64	34.5%	60	4
dMyc	151	24	15.8%	24	0
dMax	365	56	15.3%	42	14
dMnt	131	44	33.5%	43	1
Total targets	968	253	26.1%	232	21

Column A subgroups: targets were sorted according to subgroup specification as reflected in Figure 2F. Column B: Number of loci detected as statistically significant *DamID* targets. Column C: Number of genes that are depicted in column B that also show dMyc-regulated gene expression changes.

dMnt (60%) and by all three Max network proteins (48%; Table 3). Furthermore, applying the REDUCE algorithm to the gene expression profile revealed a significant correlation with E-box and DRE sequences, supporting the notion that our binding loci correspond to transcriptionally regulated genes (see Supplemental Material S5). Within the group of genes bound by both dMyc and dMnt no genes were found to be repressed by dMyc, reinforcing the notion that the distinct biological effects of dMyc and dMnt derives from their opposing transcriptional functions (see below). Of the total dMax network targets identified by *DamID* and whose expression is dMyc regulated (253), the majority (232) were activated and only 21 repressed. Thus, the other 280 repressed genes are likely to be due to indirect or downstream effects of dMyc.

Chromatin immunoprecipitation analysis of dMyc and dMnt binding and histone acetylation at the *bic* locus

We employed chromatin immunoprecipitation (ChIP) to assess whether dMyc and dMnt bind the same site proximal to a *DamID*-identified target gene and alter the status of histone tail acetylation. We chose *Bicaudal* (*bic*), as an example of a gene that displayed binding by dMyc, dMax, and dMnt (Fig. 1, *bic* is circled) and whose mRNA is also induced by dMyc in our expression analysis. *bic* is related to vertebrate BTF3, a general transcription factor whose expression was found to correlate with c-Myc in a global expression study (Zheng et al. 1987; Watson et al. 2002). *bic* possesses one E-box in the vicinity of its transcriptional start site (Fig. 4A). Using the ChIP assay we specifically detect dMyc and dMnt association with the *bic* E-box region in cells overexpressing each protein (Fig. 4B, cf. IgG control lanes 3,4 and dMyc binding lanes 5,6, and dMnt binding lanes 7,8). Importantly, binding by dMyc and dMnt alters the status of endogenous histone acetylation at this site: dMyc binding correlates with hyperacetylation of both histones H3 and H4 (Fig. 4B, lanes 11,13), whereas dMnt binding correlates with deacetylation (Fig. 4, lanes 12,14). These findings support the notion that dMyc and dMnt are capable of alternate occupation of the same binding site, as predicted from

the *DamID* data, and promote opposing changes in chromatin modification.

To determine whether the endogenous dMnt protein occupies the *bic* locus E-box, we induced *Kc* cell differentiation using ecdysone. Other work has shown that dMnt is induced during differentiation and by ecdysone in pupal stages (L.W.M. Loo and C. Thummel, unpubl.). Figure 4C shows that endogenous dMnt associates with the *bic* E-box only following ecdysone treatment (Fig. 4C, cf. lanes 6 and 7). Similar experiments with endogenous dMyc were not feasible because of the inability of our antibody to detect the low levels of endogenous protein. Importantly, dMnt binding correlates with deacetylation of histone H3 and H4 (Fig. 4C, lanes 8–12), consistent with the ability of dMnt to recruit a mSin3–HDAC complex to its binding site (L.W.M. Loo and R.N. Eisenman, in prep.).

Discussion

A significant gap in our understanding of the function of many transcriptional regulatory proteins has been the lack of comprehensive identification of their in vivo binding sites and the genes whose expression they regulate. This problem is especially pertinent for transcription factors such as Myc, Mad/Mnt, Max, and other members of the Max network that function as relatively weak transcriptional regulators, whose consensus binding site is ubiquitous, and whose expression elicits profound effects on cell growth and proliferation. Standard methods of target gene evaluation do not reliably differentiate between genes bound and directly regulated by Myc and Mad from genes whose expression is altered as a secondary or later consequence of Myc or Mad induction. In principle, it is important to know about both sets of genes, but it is also crucial to distinguish between them. The *DamID* method employed in this paper permits determination of transcription factor binding site regions in live cells and is not dependent on chemical cross-linkers or PCR primers. Because it involves “marking” of DNA in chromatin by a methyltransferase linked to a transcription factor, even transient or low affinity interactions with DNA, as well as proximity to regions

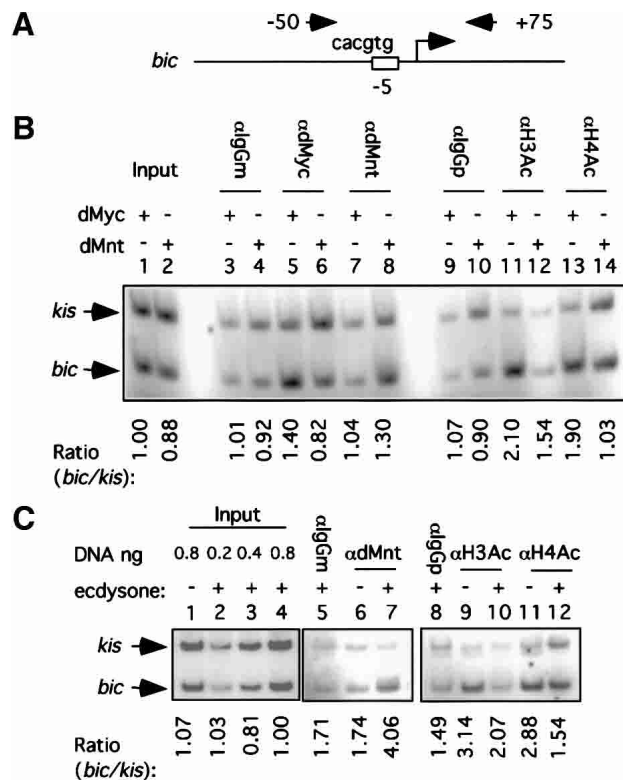


Figure 4 Occupancy of the *bic* promoter by dMyc and dMnt proteins correlates with changes in histone acetylation. (A) Diagram of the *bic* promoter, the CACGTG E-box is denoted as an open box. The PCR primer set (-50 left primer, +75 right primer) used is denoted by upper black arrows. (B) Promoter occupancy was determined using monoclonal antibodies to dMyc (lanes 5,6), dMnt (lanes 7,8), acetylated histone H3 (lanes 11,12), and acetylated histone H4 (lanes 13,14). ChIP using genomic extracts derived from dMyc-expressing cells (odd-numbered lanes) or dMnt-expressing cells (even-numbered lanes) is described in Materials and Methods. Input, 2 ng non-IP genomic DNA; *bic*, the *bic* E-box specific primers; *kis*, control primers for the *kis* gene promoter serving as a loading control; α IgGp and α IgGm, Anti-mouse polyclonal and monoclonal IgG antibodies. Enrichment is represented as the ratio between the *bic* and the *kis* PCR products as described in Materials and Methods. (C) Ecdysone treatment leads to endogenous dMnt recruitment to the *bic* promoter and correlates with deacetylation of histone H3 and H4. Ecdysone treatment and ChIP were as described in Materials and Methods. Amplification linearity was tested in the presence of ecdysone (lanes 2–4). All other abbreviations are as depicted in Figure 4B.

distal to the binding site (through looping or higher-order folding), might be detected (van Steensel and Henikoff 2000; van Steensel et al. 2001). Because we used a cDNA array to detect targeted methylation regions, only binding sites within a few kb of transcription units are detected. Therefore, our enumeration of dMax network binding sites is likely to be an underestimate. The mapping resolution also does not permit precise pinpointing of the binding site within each probed locus, although the REDUCE analysis strongly suggests that E-box motifs within target loci mediate the protein recruitment (e.g., as for *bic*; Fig. 4).

The validity of our approach is strongly supported by several lines of evidence. First, the degree of overlap between dMyc, dMax, and dMnt binding regions (Fig. 2) is consistent with the relationship between E-box binding and heterodimerization with Max established previously for the vertebrate proteins as well as for their orthologs in *Drosophila* (Blackwood and Eisenman 1991; Ayer et al. 1993; Zervos et al. 1993; Gallant et al. 1996; L.W.M. Loo and R.N. Eisenman, in prep.). Importantly, the GAGA factor, a ubiquitous transcription factor unrelated to the dMax network, displays only minimal overlap with dMnt binding sites (Fig. 1F), suggesting our results are specific for binding by dMax network transcription factors. Furthermore, studies in mammalian cells have shown both overlapping and nonoverlapping functions and target genes for Myc and Mad proteins (O'Hagan et al. 2000; Iritani et al. 2002; James and Eisenman 2002) in agreement with our DamID findings. Second, using a ChIP assay, the direct binding of dMyc and dMnt to a DamID-defined target gene, *bic* (*bicaudal*), was demonstrated (Fig. 4). In addition, the mammalian orthologs of at least 18 genes identified as binding targets for dMyc, dMax, and dMnt in our study have been demonstrated to be direct targets for vertebrate Myc using ChIP (Fernandez et al. 2003) and are indicated in Figure 3. Third, application of the REDUCE algorithm, which correlates binding with the occurrence of DNA sequence motifs (Bussemaker et al. 2001) reveals a statistically significant correlation between the E-box CACGTG and the presence of dMnt binding regions. CACGTG enrichment also correlated with dMyc binding in the presence of high dMax levels, for dMax binding in the presence of high dMyc levels, and for genes whose expression is modulated by dMyc. Fourth, a substantial set of target genes identified in our *Drosophila* gene expression microarray analysis, employing larvae overexpressing dMyc, correspond to target genes defined by DamID (Fig. 3; Table 3; Supplemental Material S1). In addition, target genes identified here are in accord with genes regulated by Myc and Mad as described in several recently published gene-expression studies in vertebrate systems (Coller et al. 2000; Guo et al. 2000; Boon et al. 2001; Neiman et al. 2001; Schuhmacher et al. 2001; Iritani et al. 2002).

Drosophila Max network binding sites define biological functions of the network

We have used the *Drosophila* Gene Ontology Database to derive an unbiased classification of genes associated with dMax network binding regions (see Table 2 for a partial list; for complete list, see Supplemental Material S2). Many of our dMax network targets identified are genes that fit well with the established biological functions of Myc and Mad. In addition, a significant number of targets point to new pathways likely to be regulated by the network. Our data demonstrate both binding to, and regulation of, genes encoding proteins broadly involved in biosynthetic processes, in accord with genetic and biochemical analyses demonstrating that Myc is involved

in cell growth in *Drosophila* and vertebrates (Schmidt 1999; Schuhmacher et al. 1999; Kim et al. 2000; de Alboran et al. 2001; Douglas et al. 2001), and from earlier global gene expression studies (see <http://www.myc-cancer-gene.org/index.asp> for a compilation of candidate target genes; Guo et al. 2000; Boon et al. 2001; Neiman et al. 2001; Schuhmacher et al. 2001). Our DamID binding loci also include genes involved in cell cycle and DNA replication (Table 2). Our list of putative dMax network targets also reveals potential novel pathways such as mitochondrial biogenesis and function, as well as vesicular transport. Other pathways known to be linked to Myc such as apoptosis, proteolysis, and the immune response are also reflected in our list of dMax network target genes as are a number of transcription factors (Table 2).

Widespread genomic binding by the dMax network

Our findings demonstrate a surprisingly large number (968) of binding sites for proteins of the dMax network (Fig. 2). Considering that the array represents a random sampling of ~50% of *Drosophila* coding regions, a conservative estimate is that dMax network proteins interact with ~2000 genes, and, as mentioned above, this is likely to be an underestimate. It is important to note however that dMax network proteins do not bind profiligately to DNA, as evidenced by the low degree of overlap with GAGA factor, the general correlation of E-box sequences with binding, and the lack of association with repeat elements linked to HP1 binding previously. HP1 is predominantly localized to pericentric heterochromatin, and its binding is associated with silenced chromatin structure (James et al. 1989). The lack of association of dMyc, dMax, or dMad with such elements may indicate that the network proteins are primarily associated with genes that are subject to ongoing transcriptional modulation. These findings are in accord with extensive ChIP assays carried out by Fernandez et al. (2003) in human cells. That study suggests that 8%–10% of cellular genes associate with Myc and in general display enhanced histone H3 and H4 acetylation.

The large number of binding sites and regulated target genes identified in this study contrasts with earlier ideas of Myc function that posited a small number of critical targets. However, not all binding sites necessarily result in direct transcriptional regulation by dMax network factors. This is evident from our dMyc-dependent gene expression data carried out in growing third instar larvae. At this developmental stage, 31% (89/287) of our Myc binding loci (as determined in *Kc* cells) displayed altered mRNA expression in larvae. Of genes that were detected as overlapping targets of all three proteins or of only dMyc and dMnt, 48.6% and 60.5% respectively, displayed concomitant changes in mRNA levels upon Myc induction (Fig. 3; Table 3). Interestingly, Fernandez et al. (2003) describe Myc binding and histone acetylation at mammalian genes whose expression does not appear to change in response to induction of Myc. One possible explanation is that Myc binding to a subset of genes, although not immediately affecting gene expression,

confers a permissive state on chromatin allowing binding by other cis-acting factors at later times.

The many dMax targets detected that are shared with dMyc and dMnt most likely represent binding by dMyc–dMax and dMnt–dMax heterodimers. However, the extent of nonoverlap between binding sites for these proteins is more extensive than expected. For example, we found that dMax expressed at low levels binds to 365 genes that do not overlap with either dMnt or dMyc targets. However, 15% of these binding loci are regulated by dMyc in our larval expression analysis. Thus, the degree of overlap is probably influenced by the temporal pattern and levels of dMyc expression. This has implications for tumorigenesis where vertebrate Myc proteins are often dramatically overexpressed. Our work provides evidence that such overexpression may shift the spectrum of target genes relative to those expressed in normal cells.

Max homodimers bind E-boxes with relatively low affinity and in mammalian cells are inhibited by phosphorylation from binding DNA (Berberich and Cole 1992; Koskinen et al. 1994). Although we do not know whether dMax homodimers are similarly blocked from binding to DNA in vivo, we favor the idea that the large number (365) of unique dMax binding sites and the lack of correlation with E-boxes reflects dimerization and DNA binding by dMax with as-yet-unidentified interacting proteins. Interestingly, in mammalian cells Max has been found, in association with the bHLHZ protein Mga, in E2F6 repression complexes (Ogawa et al. 2002). Similarly, unique sites found for dMnt and dMyc may represent non-E-box DNA binding through formation of higher-order complexes. For mammalian Myc, interaction of Myc–Max heterodimers with the Miz-1 protein has been shown to direct Myc to non-E-box sites (Staller et al. 2001). It is likely that associations with other partners may redirect dMyc and dMnt to unique binding sites. If so, our findings indicate that such interactions may be extensive and are an important part of dMax network function.

The canonical E-box sequence alone is unlikely to be sufficient to determine specific binding by dMax network proteins and, indeed, many E-box-containing promoters are not associated with Max network proteins (Bouchard et al. 2001; Fernandez et al. 2003). One possibility is that other sequences in the vicinity of an E-box may play a role in target gene specificity. For example, the DRE, which correlated with binding of all three dMax network proteins is located within <1 kb of many of our E-box sequences. Therefore, it is tempting to hypothesize that the DRE operates in cis with adjacent E-boxes to recruit protein complexes that will either promote activation or repression. Alternatively, the proximity of DRE and E-box sites may reflect coordinate regulation of the same genes through distinct signaling pathways.

In addition, REDUCE analysis has revealed a number of unexpected correlations. For example, we found association between dMyc and AT-rich sequences when dMax levels are limiting (Supplemental Material S5). In several loci examined, these AT-rich regions occur in the

vicinity of genes lacking E-boxes, perhaps reflecting dMyc association with as-yet-undefined binding proteins when dMax levels are limiting. REDUCE analysis of dMax binding regions failed to detect a binding correlation with CACGTG. However, when high levels of dMyc were expressed together with dMax–Dam, REDUCE analysis of dMax binding regions found the E-box significantly correlated with binding (data not shown). This is in accord with previous data that dMax homodimers bind only weakly to E-boxes, and that Max binding is largely directed by its heterodimeric partners (Berberich and Cole 1992; Fieber et al. 2001). Perhaps the AT- (and CG-) rich sequences influence architecture of the binding site or serve as binding motifs for factors that enhance dMax network protein association with DNA.

Taken together, our data suggest a rather more complex picture of the functioning of Max network transcription factors than has been considered previously. The results suggest extensive yet specific interaction with chromatin probably encompassing thousands of binding sites and directly affecting expression of hundreds of genes. In addition, the DamID results indicate the possibility of several different modes of Myc, Max, and Mad/Mnt interactions. These include binding to partner proteins yet to be identified as well as potential cooperation with other transcription factors. Earlier experiments have shown that Myc and Mad expression is under tight control by the cell. Such control is likely to be important in balancing the multiple protein–protein and DNA binding interactions inferred from our data.

Materials and methods

Plasmids and constructs

Generation of chimeric Dam–proteins: Full-length cDNAs encoding *Drosophila melanogaster* Max, Mnt, Myc, and Myc Δ ct, (amino acids 1–624) were cloned in either pNDam-9E10 or p9E10-CDam vectors (van Steensel and Henikoff 2000) resulting in the formation of either N-terminal or C-terminal Dam-fusion proteins. GAF-Dam was described previously (van Steensel et al. 2001). For expression in *Kc* cells, the appropriate cDNAs were subcloned into heat shock (hs)-inducible pCasper vector or pMTV, a copper-inducible insect expression vector (Invitrogen). All constructs were verified by automated sequencing.

Cell culture, transfections, and reporter assays

Reporter assays were performed as described previously using pCI-Neo to express the different proteins in 293T cells (Laherty et al. 1997). Kc167 *Drosophila* cells were maintained, transfected, and stained as described earlier (van Steensel et al. 1995; Henikoff et al. 2000). Cells were treated with 0.2 μ M 20-hydroxy ecdysone (Sigma) for 18 h where indicated. Cell cultures for microarray analysis were as follows: Where indicated, pMTV–dMyc (Cu²⁺-inducible) was cotransfected in the presence of 0.5 mM CuSO₄ (Sigma), along with hs-dMax–Dam. Alternatively, cells were transfected with a plasmid coding for either hs-dMax (low Max) or copper-induced pMTV–dMax (high Max), or pMTV–dMax Δ Zip (Δ 92–115) along with hs-Ndam–Myc. pMTV–dMax was also cotransfected with hs-dMnt–Dam. All assays were carried out in the absence of heat shock.

Microarray analysis and chromatin profiling

All chromatin profiles were performed as described (van Steensel et al. 2001) using spotted microarrays constructed from release 1 of the *Drosophila* Gene Collection (Rubin et al. 2000) and 430 additional cDNA and genomic sequences. Three independent experiments were performed for each of the proteins. In an additional experimental set the dye labeling was reversed for Dam and the fusion protein to exclude bias related to the dye. Microarray images were quantified using GenePix Pro v3.0 imaging software (Axon Instruments). Microarray results were analyzed using Cyber-T microarray analysis software (Baldi and Long 2001) and corresponding Bayesian-derived P values were adjusted for multiple hypotheses testing using a false discovery rate (FDR) method (Benjamini and Hochberg 1995) where FDR was set at 0.05. Additionally, dMnt–Dam versus dMnt–Dam control experiments were used to assess experimental array variation within the DamID assay. Accordingly, a lower-bound ratio threshold of log₂ 0.58 (i.e., 3 \times S.D.) was also imposed. For access to DamID primary data, see Supplemental Material S3.

Larval dMyc expression profile experiments

Fly strains: All transgenes are P[+] in w⁺ strains. w⁺;+;Act 5c > CD2 > GAL4 UAS–GFP (Neufeld et al. 1998); y w hs-FLP¹²²; +; UAS–dMyc (Zaffran et al. 1998). y w hs-FLP¹²²; +; +. Adult flies and larvae were raised in regular fly food consisting of cornmeal and molasses at 25°C. Larvae overexpressing either UAS-regulated dMyc;GFP or GFP alone transgenes were generated using the Flp/Gal4 method (Struhl and Basler 1993; Pignoni and Zipursky 1997; Neufeld et al. 1998). Larvae were staged from hatching and raised at a density of 50 per vial at 25°C. Third instar larvae (110 h after egg deposition, AED) were heat shocked at 37°C for 2 h, and larvae were collected 7 h after heat shock (~120 h AED). Total RNA was isolated using TRIzol reagent (Invitrogen) as described by manufacturer followed by RNeasy (Qiagen) clean up. cRNA targets were generated using a standard amino-allyl labeling protocol, where 30 μ g each of “experimental” (dMyc;GFP: hs-FLP¹²²; Act–GAL4, UAS–GFP; UAS–dMyc) and “reference” (GFP only :yw hs-FLP¹²²; Act–GAL4, UAS–GFP; +) total RNAs were coupled to either Cy3 or Cy5 fluorophores. Paired labeled targets were processed on microarrays using protocols described elsewhere (Fazzio et al. 2001). Posthybridized arrays were scanned using a GenePix 4000 scanner (Axon Instruments). Data were generated from five independent replicates (two with one dye orientation and three with the reversed dye orientation) at 7 h and four independent replicates (two with one dye orientation and two with the reversed dye orientation) at 14 h. Spot-level ratios were log₂ transformed and a loess normalization ($f = 0.33$) strategy was applied using S-Plus (MathSoft, Cambridge, MA) to correct for observed intra-array intensity-dependent ratio biasing. Expression data analysis was performed as described for DamID. Accordingly, any gene/est with a P value that meets the FDR = 0.05 criterion and has a fold-change outside of the range ± 1.5 were identified as differentially expressed. Access to gene expression data is available as Supplemental Material (S4).

Motif-based regression analysis of Max network proteins binding using REDUCE

To identify motifs responsible for the preferential binding of the transcription factors studied, we used REDUCE, a tool originally developed for the analysis of gene expression data but which can also be used to analyze binding data. To remove any remaining bias attributable to the varying number of matches with GATC (the restriction site used in the DamID procedure)

in the probe sequences, we first fitted a linear model based on GATC. The corrected log-ratios were subsequently analyzed as described [Bussemaker et al. 2001] using software available at <http://bussemaker.bio.columbia.edu/reduce>. REDUCE relies on standard linear regression of the binding log-ratio for each probe on the number of matches of a given motif to the DNA sequence associated with the binding log-ratio (i.e., the probed region plus flanking sequence).

The regression coefficient (slope) associated with each motif is determined by performing a least-squares fit of this model to the data; it estimates the contribution to binding to a locus by each occurrence of the motif. The value of R^2 for this fit is equal to the square of the Pearson correlation r between the binding log-ratio and the motif count; it measures the statistical significance of the regression coefficient. The P value associated with R^2 has to be corrected to control the family-wise error rate in the parallel test of a large number of motifs. We used the Bonferroni correction, which amounts to multiplying the P value by the total number of motifs tested. The complete REDUCE search results can be found in Supplemental Material S5.

Chromatin immunoprecipitation

Chromatin fixation and purification procedures were as described [Schubeler et al. 2000; Takahashi et al. 2000; Sawado et al. 2001]. Where indicated, cells were treated with 0.5 mM CuSO_4 . Immunoprecipitation of cross-linked DNA was carried out using the following antibodies: rabbit polyclonal αAcH_3 , αAcH_4 (Upstate), αdMnt mouse monoclonal antibodies (b8, b11; L. Loo, in prep.), and αdMyc mouse monoclonal (b10; Prober and Edgar 2002). Duplex ^{32}P -dCTP-labeled PCR reactions were preformed and quantified using "Storm" PhosphorImager system (Molecular Dynamics). Specific protein binding was measured as a comparison between the relative ratio of intensity of the specific target "bic" PCR products to the reference. The reference gene *kismet* (*kis*) had negative binding values for both dMyc and dMnt, was not detected as a target in the larval myc expression experiment, had linear amplification profile under the various conditions, and, therefore, serves as a loading control. Primer sequences can be found in Supplemental Material.

Acknowledgments

We are grateful to Steve Henikoff, Dirk Schübeler, Harmit Malik, Fumiko Hirose, and Masamitsu Yamaguchi for valuable reagents, advice, and unpublished data. We are also grateful to Bruno Amati and Paula Fernandez for discussion and sharing unpublished data. We thank Steve Henikoff and Carla Grandori for critical readings of the manuscript and Grace Navaja for technical assistance. Support for this work was from the Human Frontiers Science Program (A.O.), NIH grants GM47852 (S.P.), CA80295-03S (E.W.), 1P20LM007276-01 (HJB), Uehara Memorial Foundation (T.S.), RO1CA57138 (R.N.E.), and NIH/NCI Cancer Center Support Grant P30 CA15704-29 (J.D.), and R01 GM51186 (B.A.E.).

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

References

Amati, B., Frank, S.R., Donjerkovic, D., and Taubert, S. 2001. Function of the c-Myc oncoprotein in chromatin remodeling

- and transcription. *Biochim. Biophys. Acta* **1471**: M135–M145.
- Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. 2000. Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**: 25–29.
- Ayer, D.E., Kretzner, L., and Eisenman, R.N. 1993. Mad: A heterodimeric partner for Max that antagonizes Myc transcriptional activity. *Cell* **72**: 211–222.
- Baldi, P. and Long, A.D. 2001. A Bayesian framework for the analysis of microarray expression data: Regularized t-test and statistical inferences of gene changes. *Bioinformatics* **17**: 509–519.
- Benjamini, Y. and Hochberg, Y. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J.R. Statistical Soc. Ser. B-Methodological* **57**: 289–300.
- Berberich, S.J. and Cole, M.D. 1992. Casein kinase II inhibits the DNA-binding activity of Max homodimers but not Myc/Max heterodimers. *Genes & Dev.* **6**: 166–176.
- Blackwell, T.K., Kretzner, L., Blackwood, E.M., Eisenman, R.N., and Weintraub, H. 1990. Sequence-specific DNA-binding by the c-Myc protein. *Science* **250**: 1149–1151.
- Blackwell, T.K., Huang, J., Ma, A., Kretzner, L., Alt, F.W., Eisenman, R.N., and Weintraub, H. 1993. Binding of Myc proteins to canonical and noncanonical DNA sequences. *Mol. Cell. Biol.* **13**: 5216–5224.
- Blackwood, E.M. and Eisenman, R.N. 1991. Max: A helix-loop-helix zipper protein that forms a sequence-specific DNA binding complex with Myc. *Science* **251**: 1211–1217.
- Boon, K., Caron, H.N., van Asperen, R., Valentijn, L., Hermus, M.-C., van Sluis, P., Roobeek, I., Weis, I., Voute, P.A., Schwab, M., et al. 2001. N-myc enhances the expression of a large set of genes functioning in ribosome biogenesis and protein synthesis. *EMBO J.* **20**: 1383–1393.
- Bouchard, C., Dittrich, O., Kiermaier, A., Dohmann, K., Menkel, A., Eilers, M., and Luscher, B. 2001. Regulation of cyclin D2 gene expression by the Myc/Max/Mad network: Myc-dependent TRRAP recruitment and histone acetylation at the cyclin D2 promoter. *Genes & Dev.* **15**: 2042–2047.
- Bourbon, H.M., Gonzy-Treboul, G., Peronnet, F., Alin, M.F., Ardourel, C., Benassayag, C., Cribbs, D., Deutsch, J., Ferrer, P., Haenlin, M., et al. 2002. A P-insertion screen identifying novel X-linked essential genes in *Drosophila*. *Mech. Dev.* **110**: 71–83.
- Bussemaker, H.J., Li, H., and Siggia, E.D. 2001. Regulatory element detection using correlation with expression. *Nat. Genet.* **27**: 167–171.
- Coller, H.A., Grandori, C., Tamayo, P., Colbert, T., Lander, E.S., Eisenman, R.N., and Golub, T.R. 2000. Expression analysis with oligonucleotide microarrays reveals MYC regulates genes involved in growth, cell cycle, signaling, and adhesion. *Proc. Natl. Acad. Sci.* **97**: 3260–3265.
- Davis, A.C., Wims, M., Spotts, G.D., Hann, S.R., and Bradley, A. 1993. A null c-myc mutation causes lethality before 10.5 days of gestation in homozygous and reduced fertility in heterozygous female mice. *Genes & Dev.* **7**: 671–682.
- de Alboran, I.M., O'Hagan, R.C., Gartner, F., Malynn, B., Davidson, L., Rickert, R., Rajewsky, K., DePinho, R.A., and Alt, F.W. 2001. Analysis of C-MYC function in normal cells via conditional gene-targeted mutation. *Immunity* **14**: 45–55.
- Douglas, N.C., Jacobs, H., Bothwell, A.L., and Hayday, A.C. 2001. Defining the specific physiological requirements for c-Myc in T cell development. *Nat. Immunol.* **2**: 307–315.
- Eisenman, R.N. 2001. Deconstructing myc. *Genes & Dev.*

- 15: 2023–2030.
- Fazio, T.G., Kooperberg, C., Goldmark, J.P., Neal, C., Basom, R., Delrow, J., and Tsukiyama, T. 2001. Widespread collaboration of Isw2 and Sin3-Rpd3 chromatin remodeling complexes in transcriptional repression. *Mol. Cell. Biol.* **21**: 6450–6460.
- Fernandez, P.C., Frank, S.R., Wang, L., Schroeder, M., Liu, S., Greene, J., Cocito, A., and Amati, B. 2003. Genomic targets of the human c-Myc protein. *Genes & Dev.* (this issue).
- Fieber, W., Schneider, M.L., Matt, T., Krautler, B., Konrat, R., and Bister, K. 2001. Structure, function, and dynamics of the dimerization and dna-binding domain of oncogenic transcription factor v-myc. *J. Mol. Biol.* **307**: 1395–1410.
- Frank, S.R., Schroeder, M., Fernandez, P., Taubert, S., and Amati, B. 2001. Binding of c-Myc to chromatin mediates mitogen-induced acetylation of histone H4 and gene activation. *Genes & Dev.* **15**: 2069.
- Gallant, P., Shiio, Y., Cheng, P.F., Parkhurst, S., and Eisenman, R.N. 1996. Myc and Max homologs in *Drosophila*. *Science* **274**: 1523–1527.
- Grandori, C., Cowley, S.M., James, L.P., and Eisenman, R.N. 2000. The MYC/MAX/MAD network and the transcriptional control of cell behavior. *Annu. Rev. Cell Dev. Biol.* **16**: 653–699.
- Guo, Q.M., Malek, R.L., Kim, S., Chiao, C., He, M., Ruffy, M., Sanka, K., Lee, N.H., Dang, C.V., and Liu, E.T. 2000. Identification of c-myc responsive genes using rat cDNA microarray. *Cancer Res.* **60**: 5922–5928.
- Henikoff, S., Ahmad, K., Platero, J.S., and van Steensel, B. 2000. Heterochromatic deposition of centromeric histone H3-like proteins. *Proc. Natl. Acad. Sci.* **97**: 716–721.
- Hirose, F., Yamaguchi, M., Handa, H., Inomata, Y., and Matsukage, A. 1993. Novel 8-base pair sequence (*Drosophila* DNA replication-related element) and specific binding factor involved in the expression of *Drosophila* genes for DNA polymerase alpha and proliferating cell nuclear antigen. *J. Biol. Chem.* **268**: 2092–2099.
- Hirose, F., Ohshima, N., Shiraki, M., Inoue, Y.H., Taguchi, O., Nishi, Y., Matsukage, A., and Yamaguchi, M. 2001. Ectopic expression of DREF induces DNA synthesis, apoptosis, and unusual morphogenesis in the *Drosophila* eye imaginal disc: Possible interaction with Polycomb and trithorax group proteins. *Mol. Cell. Biol.* **21**: 7231–7242.
- Hochheimer, A., Zhou, S., Zheng, S., Holmes, M.C., and Tjian, R. 2002. TRF2 associates with DREF and directs promoter-selective gene expression in *Drosophila*. *Nature* **420**: 439–445.
- Iritani, B.M., Delrow, J., Grandori, C., Gomez, I., Klacking, M., Carlos, L.S., and Eisenman, R.N. 2002. Modulation of T lymphocyte development, growth, and cell size by the Myc-antagonist Mad1 transcriptional repressor. *EMBO J.* **21**: 4820–4830.
- Jackson, S.M. and Blochlinger, K. 1997. cut interacts with Notch and protein kinase A to regulate egg chamber formation and to maintain germline cyst integrity during *Drosophila* oogenesis. *Development* **124**: 3663–3672.
- James, L. and Eisenman, R.N. 2002. Myc and Mad bHLHZ domains possess identical DNA-binding specificities but only partially overlapping functions in vivo. *Proc. Natl. Acad. Sci.* **99**: 10429–10434.
- James, T.C., Eissenberg, J.C., Craig, C., Dietrich, V., Hobson, A., and Elgin, S.C. 1989. Distribution patterns of HP1, a heterochromatin-associated nonhistone chromosomal protein of *Drosophila*. *Eur. J. Cell Biol.* **50**: 170–180.
- Jasper, H., Benes, V., Atzberger, A., Sauer, S., Ansorge, W., and Bohmann, D. 2002. A genomic switch at the transition from cell proliferation to terminal differentiation in the *Drosophila* eye. *Dev. Cell* **3**: 511–521.
- Johnston, L.A., Prober, D.A., Edgar, B.A., Eisenman, R.N., and Gallant, P. 1999. *Drosophila* myc regulates growth during development. *Cell* **98**: 779–790.
- Kim, S., Li, Q., Dang, C.V., and Lee, L.A. 2000. Induction of ribosomal genes and hepatocyte hypertrophy by adenovirus-mediated expression of c-Myc in vivo. *Proc. Natl. Acad. Sci.* **97**: 11198–11202.
- Knoepfler, P.S., Cheng, P.F., and Eisenman, R.N. 2002. N-myc is essential during neurogenesis for the rapid expansion of progenitor cell populations and the inhibition of neuronal differentiation. *Genes & Dev.* **16**: 2699–2712.
- Koskinen, P.J., Västriik, I., Mäkelä, T.P., Eisenman, R.N., and Alitalo, K. 1994. Max activity is affected by phosphorylation at two NH₂-terminal sites. *Cell Growth Diff.* **5**: 313–320.
- Laherty, C.D., Yang, W.-M., Sun, J.-M., Davie, J.R., Seto E., and Eisenman, R.N. 1997. Histone deacetylases associated with the mSin3 corepressor mediate Mad transcriptional repression. *Cell* **89**: 349–356.
- Li, H. 2002. Computational approaches to identifying transcription factor binding sites in the yeast genome. *Methods Enzymol.* **350**: 484–495.
- McMahon, S.B., Wood, M.A., and Cole, M.D. 2000. The essential cofactor TRRAP recruits the histone acetyltransferase hGCN5 to c-Myc. *Mol. Cell. Biol.* **20**: 556–562.
- Neiman, P.E., Ruddell, A., Jasoni, C., Loring, G., Thomas, S.J., Brandvold, K.A., Lee, R.m., Burnside, J., and Delrow, J. 2001. Analysis of gene expression during myc oncogene-induced lymphomagenesis in the bursa of Fabricius. *Proc. Natl. Acad. Sci.* **98**: 6378–6383.
- Nepveu, A. 2001. Role of the multifunctional CDP/Cut/Cux homeodomain transcription factor in regulating differentiation, cell growth and development. *Gene* **270**: 1–15.
- Neufeld, T.P., de la Cruz, A.F.A., Johnston, L.A., and Edgar, B.A. 1998. Coordination of growth and cell division in the *Drosophila* wing. *Cell* **93**: 1183–1193.
- Ogawa, H., Ishiguro, K., Gaubatz, S., Livingston, D.M., and Nakatani, Y. 2002. A complex with chromatin modifiers that occupies E2F- and Myc- responsive genes in G0 cells. *Science* **296**: 1132–1136.
- O'Hagan, R.C., Schreiber-Agus, N., Chen, K., David, G., Engelman, J.A., Schwab, R., Alland, L., Thomson, C., Ronning, D.R., Sacchettini, J.C., et al. 2000. Gene-target recognition among members of the Myc superfamily and implications for oncogenesis. *Nat. Genet.* **24**: 113–119.
- Pignoni, F. and Zipursky, S.L. 1997. Induction of *Drosophila* eye development by decapentaplegic. *Development* **124**: 271–278.
- Prendergast, G.C. and Ziff, E.B. 1991. Methylation-sensitive sequence-specific DNA binding by the c-Myc basic region. *Science* **251**: 186–189.
- Prendergast, G.C., Lawe, D., and Ziff, E.B. 1991. Association of Myn, the murine homolog of Max, with c-Myc stimulates methylation-sensitive DNA binding and Ras cotransformation. *Cell* **65**: 395–407.
- Prober, D.A. and Edgar, B.A. 2002. Interactions between Ras1, dMyc, and dPI3K signaling in the developing *Drosophila* wing. *Genes & Dev.* **16**: 2286–2299.
- Rubin, G.M., Hong, L., Brokstein, P., Evans-Holm, M., Frise, E., Stapleton, M., and Harvey, D.A. 2000. A *Drosophila* complementary DNA resource. *Science* **287**: 2222–2224.
- Sawado, T., Igarashi, K., and Groudine, M. 2001. Activation of beta-major globin gene transcription is associated with recruitment of NF-E2 to the beta-globin LCR and gene promoter. *Proc. Natl. Acad. Sci.* **98**: 10226–10231.

- Schmidt, E.V. 1999. The role of c-myc in cellular growth control. *Oncogene* **18**: 2988–2996.
- Schreiber-Agus, N., Stein, D., Chen, K., Goltz, J.S., Stevens, L., and DePinho, R.A. 1997. *Drosophila* Myc is oncogenic in mammalian cells and plays a role in the diminutive phenotype. *Proc. Natl. Acad. Sci.* **94**: 1235–1240.
- Schubeler, D., Francastel, C., Cimbora, D.M., Reik, A., Martin, D.I., and Groudine, M. 2000. Nuclear localization and histone acetylation: a pathway for chromatin opening and transcriptional activation of the human beta-globin locus. *Genes & Dev.* **14**: 940–950.
- Schuhmacher, M., Staeger, M.S., Pajic, A., Polack, A., Weidle, U.H., Bornkamm, G.W., Eick, S., and Kohlhuber, F. 1999. Control of cell growth by c-Myc in the absence of cell division. *Curr. Biol.* **9**: 1255–1258.
- Schuhmacher, M., Kohlhuber, F., Holzel, M., Kaiser, C., Burtcher, H., Jarsch, M., Bornkamm, G.W., Laux, G., Polack, A., Weidle, U.H., et al. 2001. The transcriptional program of a human B cell line in response to myc. *Nucleic Acids Res.* **29**: 397–406.
- Seoane, J., Le, H.V., and Massague, J. 2002. Myc suppression of the p21(Cip1) Cdk inhibitor influences the outcome of the p53 response to DNA damage. *Nature* **419**: 729–734.
- Solomon, D.L.C., Amati, B., and Land, H. 1993. Distinct DNA binding preferences for the c-Myc/Max and Max/Max dimers. *Nucleic Acids Res.* **21**: 5372–5376.
- Staller, P., Peukert, K., Kiermaier, A., Seoane, J., Lukas, J., Karsunky, H., Möröy, T., Bartek, J., Massague, J., Hänel, F., et al. 2001. Repression of p15^{INK4b} expression by Myc through association with Miz-1. *Nat. Cell Biol.* **3**: 392–399.
- Stanton, B.R., Perkins, A.S., Tessarollo, L., Sassoon, D.A., and Parada, L.F. 1992. Loss of N-myc function results in embryonic lethality and failure of the epithelial component of the embryo to develop. *Genes & Dev.* **6**: 2235–2247.
- Struhl, G. and Basler, K. 1993. Organizing activity of wingless protein in *Drosophila*. *Cell* **72**: 527–540.
- Takahashi, Y., Rayman, J.B., and Dynlacht, B.D. 2000. Analysis of promoter binding by the E2F and pRB families in vivo: Distinct E2F proteins mediate activation and repression. *Genes & Dev.* **14**: 804–816.
- Trumpp, A., Refaeli, Y., Oskarsson, T., Gasser, S., Murphy, M., Martin, G.R., and Bishop, J.M. 2001. c-Myc regulates mammalian body size by controlling cell number but not cell size. *Nature* **414**: 768–773.
- van Steensel, B. and Henikoff, S. 2000. Identification of in vivo DNA targets of chromatin proteins using tethered Dam methyltransferase. *Nat. Biotech.* **18**: 424–428.
- van Steensel, B., Brink, M., van der Meulen, K., van Binnendijk, E.P., Wansink, D.G., de Jong, L., de Kloet, E.R., and van Driel, R. 1995. Localization of the glucocorticoid receptor in discrete clusters in the cell nucleus. *J. Cell Sci.* **108**: 3003–3011.
- van Steensel, B., Delrow, J., and Henikoff, S. 2001. Chromatin profiling using targeted DNA adenine methyltransferase. *Nat. Genet.* **27**: 304–308.
- van Steensel, B., Delrow, J., and Bussemaker, H.J. 2003. Genome-wide analysis of *Drosophila* GAGA factor target genes reveals context-dependent DNA binding. *Proc. Natl. Acad. Sci.* **100**: 2580–2585.
- Watson, J.D., Oster, S.K., Shago, M., Khosravi, F., and Penn, L.Z. 2002. Identifying genes regulated in a Myc-dependent manner. *J. Biol. Chem.* **277**: 36921–36930.
- Zaffran, S., Chartier, A., Gallant, P., Astier, M., Arquier, N., Doherty, D., Gratecos, D., and Semeriva, M. 1998. A *Drosophila* RNA helicase gene, *pitchoune*, is required for cell growth and proliferation and is a potential target of d-Myc. *Development* **125**: 3571–3584.
- Zervos, A.S., Gyuris, J., and Brent, R. 1993. Mx1, a protein that specifically interacts with Max to bind Myc–Max recognition sites. *Cell* **72**: 223–232.
- Zhao, K., Hart, C.M., and Laemmli, U.K. 1995. Visualization of chromosomal domains with boundary element-associated factor BEAF-32. *Cell* **81**: 879–889.
- Zheng, X.M., Moncollin, V., Egly, J.M., and Chambon, P. 1987. A general transcription factor forms a stable complex with RNA polymerase B (II). *Cell* **50**: 361–368.
- Zhou, Z.Q. and Hurlin, P.J. 2001. The interplay between Mad and Myc in proliferation and differentiation. *Trends Cell Biol.* **11**: S10–14.
- Ziegelbauer, J., Shan, B., Yager, D., Larabell, C., Hoffmann, B., and Tjian, R. 2001. Transcription factor MIZ-1 is regulated via microtubule association. *Mol. Cell* **8**: 339–349.